Journal of Electromyography and Kinesiology 25 (2015) 479–487

Contents lists available at ScienceDirect

Journal of Electromyography and Kinesiology

journal homepage: www.elsevier.com/locate/jelekin

A proposed method to detect kinematic differences between and within individuals

David M. Frost [a], Tyson A.C. Beach [a], Stuart M. McGill [b], Jack P. Callaghan [b,*]

[a] Faculty of Kinesiology and Physical Education, University of Toronto, 55 Harbord Street, Toronto, Ontario M5S 2W6, Canada
[b] Department of Kinesiology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

ARTICLE INFO

Article history:
Received 7 September 2014
Received in revised form 20 February 2015
Accepted 27 February 2015

Keywords:
Firefighter
Movement variability
Injury
Prevention
Single-subject

ABSTRACT

The primary objective was to examine the utility of a novel method of detecting "actual" kinematic changes using the within-subject variation. Twenty firefighters were assigned to one of two groups (lifting or firefighting). Participants performed 25 repetitions of two lifting or firefighting tasks, in three sessions. The magnitude and within-subject variation of several discrete kinematic measures were computed. Sequential averages of each variable were used to derive a cubic, quadratic and linear regression equation. The efficacy of each equation was examined by contrasting participants' sequential means to their 25-trial mean ± 1 SD and 2 SD. The magnitude and within-subject variation of each dependent measure was repeatable for all tasks; however, each participant did not exhibit the same movement patterns as the group. The number of instances across all variables, tasks and testing sessions whereby the 25-trial mean ± 1 SD was contained within the boundaries established by the regression equations increased as the aggregate scores included more trials. Each equation achieved success in at least 88% of all instances when three trials were included in the sequential mean (95% with five trials). The within-subject variation may offer a means to examine participant-specific changes without having to collect a large number of trials.

1. Introduction

Because people exhibit characteristic motion patterns when executing specific tasks (Morriss et al., 1997; Rodano and Squadrone, 2002), and unique responses to changes in the task or environmental constraints (Caster and Bates, 1995; Dufek and Bates, 1990; Dufek et al., 1995), there is a need to establish the range of typical variation outside of which reflects an actual change in movement behavior. More specifically, it would be helpful to know when a kinematic difference is of a magnitude that is beyond what would be expected by the normal trial-to-trial variation. In the presence of between-subject variation (i.e. heterogeneity), averaging data across participants could inadvertently lead to the conclusion of "no significant intervention effect" when in fact substantial and clinically-relevant adaptations did occur within some individuals (James and Bates, 1997). Also possible is the finding of a statistically significant effect, but the aggregated (group-level) response misrepresented the breadth of individual strategies that were used (Caster and Bates, 1995; Dufek and Bates, 1990; Dufek et al., 1995; Scholes et al., 2012). For example, Dufek et al. (1995) examined how individuals adapted their maximum vertical ground reaction force while running in response to variations in stride length, and found that nobody exhibited the group's average strategy. If no individual reflects the mean, then the assessment of group averages compromises the ability to investigate several questions and probe different mechanisms.

Understanding the degree to which a specific pattern or descriptor of motion varies across a population can assist in developing an effective intervention; however, an effort must also be made to estimate the variation that would be expected within individuals. Hopkins (2000) suggested that intra-individual variability is important for researchers because it impacts the precision of all experimental variables, and could adversely affect any conclusions and recommendations. One approach to address this challenge in intervention research may be to identify the descriptors of motion that are least variable and thus better indicators of change, though it cannot be assured that these descriptors would be "meaningful" if selected based strictly on their variability. Alternatively, if there are specific descriptors of motion that researchers wish to change because they have been linked with adverse health (e.g. uncontrolled frontal plane knee motion (Hewett et al., 2005)) or poor

* Corresponding author. Tel.: +1 519 888 4567x37080; fax: +1 519 8850470.
E-mail address: jack.callaghan@uwaterloo.ca (J.P. Callaghan).

http://dx.doi.org/10.1016/j.jelekin.2015.02.012
1050-6411/

performance (e.g. hip extension velocity (Harbili and Alptekin, 2014)), it is more logical to first identify the magnitude of within-subject variation so that boundary criteria can be established outside of which would be considered an actual meaningful change.

In general, collecting several trials of a given task is thought to provide a more stable estimate of an individual's movement behavior (Bates et al., 1983) for the purpose of evaluating an intervention effect or contrasting multiple conditions. If too few trials are performed, the observed variation may fall inside that which is typical for the dependent measures of interest, and thus potentially mask important findings. For this reason, the minimum number of trials necessary to achieve stable estimates of various descriptors of motion have been reported for several activities, including running (Bates et al., 1983; DeVita and Bates, 1988), walking (Hamill and McNiven, 1990), vertical jumping (Rodano and Squadrone, 2002), lifting (Dunk et al., 2005), drop landing (James et al., 2007) and cricket bowling (Stuelcken and Sinclair, 2009). However, the number of trials needed has ranged between four and twenty depending on the metric of interest and activity in question. Because it is often not practical or possible to collect twenty trials of a single condition, there is a need to explore alternative solutions that can be easily integrated into a number of methodological designs while accounting for the potential within- and between-subject variation. More specifically, for the purpose of exploring how or why an individual (or group) responds to a particular intervention, it may be critical to employ an experimental approach that uses the variation displayed by each individual so that the interpretation of the effect is not limited by differences amongst the group.

Against this backdrop, the objectives of this study were two-fold: (1) to examine the within- and between-subject variation of discrete kinematic variables chosen to characterize the performance of four occupationally relevant tasks; and (2) to evaluate the potential in using the within-subject variation as a criterion with which to define within-subject differences. The kinematic variation observed within participants may provide a means to establish a range for each performer, outside of which could be defined as a genuine difference, whether 3, 10 or 25 trials of a particular task were performed, so that future work is not limited to group analyses or constrained by the heterogeneity of the participants.

2. Methods

2.1. Participant selection

Twenty firefighters (18 men and 2 women) from the Waterloo and Kitchener Fire Departments were recruited to participate in this investigation. Ten (9 men and 1 woman) were randomly assigned to each of two groups (i.e. performed general lifting or specific firefighting tasks). A description of the participants can be found in Table 1. Exclusion criteria included firefighters with known musculoskeletal injury or pain at the time of testing and those restricted to light duty work. The study was approved by the Human Research Ethics Committee of the University and all participants gave informed consent confirming their involvement, prior to beginning the study.

Table 1
The mean (standard deviation) age, height and body mass of participants in the general lifting and firefighter-specific group.

Group	Age (years)		Height (m)		Body mass (kg)	
Lifting	35.1	(7.8)	1.79	(0.03)	88.0	(13.3)
Firefighter	32.3	(6.4)	1.81	(0.07)	89.6	(16.0)

2.2. Task selection

Four occupationally relevant tasks were simulated in a laboratory setting: (1) Light box lift – from standing, a 6.8 kg box ($0.33 \times 0.33 \times 0.28$ m) was lifted from the floor to waist height and returned to the ground at a self-selected pace; (2) Heavy box lift – from standing, a 22.7 kg box ($0.33 \times 0.33 \times 0.28$ m) was lifted from the floor to waist height and returned to the ground at a self-selected pace; (3) Hose drag – a 6.4 cm diameter rope, connected to a cable machine was placed over the right shoulder and dragged forwards approximately 3 m (movement was initiated from a staggered stance with the left foot forwards); and (4) Forced entry – a ceiling-mounted "heavy bag" was struck with a 4.5 kg sledgehammer (direction of swing was self-selected). The hose drag resistance and mass of the sledgehammer were chosen to reflect the standardized protocols of the Candidate Physical Ability Test (CPAT), a test developed by the International Associations of Fire Fighters (IAFF) and Fire Chiefs (IAFC) to assess candidates' physical ability to perform tasks that are consistent with the duties of firefighting.

2.3. Experimental protocol

Participants were instrumented with infrared markers for kinematic tracking and familiarized with the tasks they would be asked to perform using a standard set of instructions. For example, they were asked to "advance forwards with the hose as they would at the scene of fire", and to "swing the sledgehammer as if forcing a door". Three to five practice trials were performed to ensure participants understood the task instructions. Individuals assigned to the general lifting task group only performed the two lifting tasks (i.e. light and heavy box lift), while participants in the specific firefighter task group performed a simulated hose drag and forced entry task. The hose drag was resisted by a load of approximately 13 kg (load attached in series with rope).

Participants in either group performed 25 repetitions (five sets of five trials) of their two assigned tasks; the 10 sets were completed in random order. Approximately 15 s and 2 min of rest were given between trials and sets, respectively. Once 5 sets of both tasks had been completed, participants were given 15 min to recover passively, and then asked to repeat the same 10-set protocol a second time (i.e. session two). Participants returned to complete a third session, identical to the first two, within one week following their initial test. No feedback was given regarding task performance at any point throughout the investigation.

2.4. Data collection and signal processing

Three-dimensional kinematic data were measured using an active optoelectronic motion capture system (Optotrak® Certus™, NDI, Waterloo, ON, Canada). The proximal and distal endpoints of the trunk, pelvis, thighs, shanks and feet were located with a digitizing probe, and the hip joint centers (HJC) and knee joint axes (KJA) were determined "functionally" using similar methods to those described by Begon et al. (2007) and Schwartz and Rozumalski (2005). Briefly, participants were asked to perform 10 repetitions of open-chain hip flexion/extension, abduction/adduction and circumduction (all with the knee extended) and open-chain knee flexion/extension for the hip and knee joint computations, respectively. Using functionally defined segment endpoints for the shank and thigh has been shown to minimize the variation introduced via digitization and thus provide a more repeatable way to create each individual's link segment model (Frost et al., 2012). Sets of 5 or 6 markers, fixed to rigid pieces of plastic, were secured to the trunk, pelvis, thighs, shanks and feet with Velcro® straps and used to track the position and orientation of each body segment within a calibrated measurement volume. One standing calibration trial was collected such that the orientation of each segment's local axis system, as defined by its endpoints, could be determined via a transformation from an axis system embedded within each rigid body. The marker data were collected at 32 Hz and smoothed using a low-pass filter (4th order, dual pass Butterworth) with a cut-off frequency of 6 Hz.

2.5. Data analyses

Participants' movement patterns were characterized by nine variables, each chosen to reflect a possible mechanism for injury (e.g. spine motion (Callaghan and McGill, 2001; Marshall and McGill, 2010)) or a coaching observation that has been cited as a possible injury risk factor (e.g. trunk angle (Marras et al., 1993; Punnett et al., 1991)). The nine variables were: (1–3) spine flexion/extension (FLX), lateral bend (BND) and axial twist (TST) – the relative orientation of the trunk was expressed with respect to the pelvis and the corresponding direction cosine matrix was decomposed with a rotation sequence of flexion/extension, lateral bend and axial twist (Cole et al., 1993) to compute the spine angle about each axis. The orientation of the lumbar spine in a relaxed upright standing trial was defined as zero degrees; (4) trunk angle relative to the vertical (TRK) – the relative orientation of the trunk (flexion/extension only) was expressed with respect to a pelvis segment that was free to move with the body but constrained about the flexion/extension axis, thereby remaining upright; (5) shank angle relative to the vertical (SHK) – the relative orientation of the left and right shank (flexion/extension only) was expressed with respect to the upright pelvis; (6) hip to ankle distance (HIP) – the position of each hip joint in the anterior/posterior (A/P) direction was described relative to the same side ankle (the upright pelvis was to define a body-fixed A/P axis); (7) knee to ankle distance (KNE) – the position of each knee joint in the A/P direction was described relative to the same side ankle; and (8–9) left (LFT) and right knee (RGT) position relative to the frontal plane – each knee joint's position (medial/lateral) was described relative to a body-fixed plane created using the corresponding hip, ankle and distal foot (Fig. 1). The SHK, HIP and KNE were only computed for the lead leg (left) of the fireground-specific tasks and defined as an average of the left and right sides for lifting.

To objectively define the start, mid-point (lifting tasks only) and end of each trial, event detection algorithms were created in Visual3D™ (Version 4, C-Motion, Inc., Germantown, MD, United States) by tracking the motion of the trunk, pelvis and whole-body center of mass. Specifically, the lifting tasks were separated into a descent and ascent phase using the vertical displacement of the center of mass, the start and end of the hose drag were described by anterior displacement of the trunk and toe-off of the forward positioned foot, respectively, and the start and end of the forced entry were defined by changes in the position of the trunk and pelvis (the end was defined as the approximate instant of contact). The forced entry task was processed to reflect a right-handed swing. Maximums, minimums, ranges and means were computed for the nine dependent variables and the "peak" of each, with the exception of BND and TST, was described as the deviation (-maximum or minimum) hypothesized to be most relevant to the types of injuries sustained by firefighters (i.e. FLX – flexion, TRK and SHK – forward bend, HIP – posterior displacement, KNE – anterior displacement, LFT and RGT – medial displacement). Peak BND and TST were described as the range (i.e. max – min) observed. The within-subject variation is presented as an aggregate score of the 25-trial standard deviations computed for each subject and session.

2.6. Statistical analyses

The 25-trial mean for each task was used to examine the between-session differences for each dependent measure. The magnitude and within-subject variation of the maximum, minimum and mean of each task were investigated separately. Between-session comparisons were made using a general linear model with one within-subject factor (IBM SPSS Statistics, Version 20.0, Armonk, NY, U.S.A.). Significant session effects were described by p-values less than 0.05. To assess the differences amongst participants, a second set of analyses was conducted whereby the participants were treated as an independent factor (i.e. blocked design). Once again, between-session comparisons were made with a general linear model, but participant was also included as a "between-subject" factor. Because the error term for the within-subject factor was equivalent to the subject $\times$ session interaction, only significant ($p < 0.05$) main effects are presented.

External load (i.e. heavy versus light lifts) and task (i.e. hose drag versus forced entry) comparisons were also made on the within-subject variation for each dependent variable. The influence of each factor (external load or task) was examined with a general linear model with one repeated measure (the data were collapsed across all three sessions), and significant differences were described by p-values less than 0.05.

2.6.1. Within-subject differences

The group's sequential mean (i.e. average of 2, 3, 4,...25 trials) and within-subject variation (i.e. each participant's between-trial standard deviation (SD)) was computed for the peak of every variable, task and testing session. To facilitate variable and task comparisons, the sequential within-subject variation was normalized by that observed over 25 trials. Based on previous work (e.g. James et al., 2007) it was hypothesized that 25 trials would be sufficient to establish a stable estimate of the group mean and between-trial variation for each individual participant, and thus also provide an opportunity to define boundary criteria outside of which would reflect an actual difference whether 3, 5, or 15 trials were collected (Fig. 2). Because the magnitude of the group mean and within-subject variation will depend on the number of trials collected, regression analyses were used to establish relationships between the upper limits of participants' kinematic variation (i.e. 25 trial mean ± 1 SD and 2 SD) and their sequential mean scores. In each instance, the difference between the sequential mean (e.g. 5-trial) and the 25-trial mean ± 1 SD and 2 SD was normalized by the sequential variation such that boundary criteria could be developed for any number of trials without knowing the true dispersion for a particular variable. The results for all variables, tasks and testing sessions were collapsed and used to create three generalizable regression equations (i.e. cubic, quadratic, linear) that could help to define limits for within-subject differences if 25 or fewer trials are collected. The most conservative upper/lower limit (i.e. that accommodating the largest variation) was used to define to the upper and lower limits for all subsequent analyses.

Using the regression equations developed, and each participant's sequential mean scores, the number of trials used to compute the sequential mean, and the group's sequential within-subject variation, limits (upper and lower) were created for each variable outside of which would be defined as different. For example, using the equation $y = (-0.05x + 3.5)z$, where y is the upper limit, x is the number of trials, and z is the group's sequential within-subject variation, differences in spine flexion would be defined by a magnitude greater than 9° if the sequential variation associated with 10 trials was 3°. The magnitude of the regression coefficients used to define the upper and lower limits were kept the

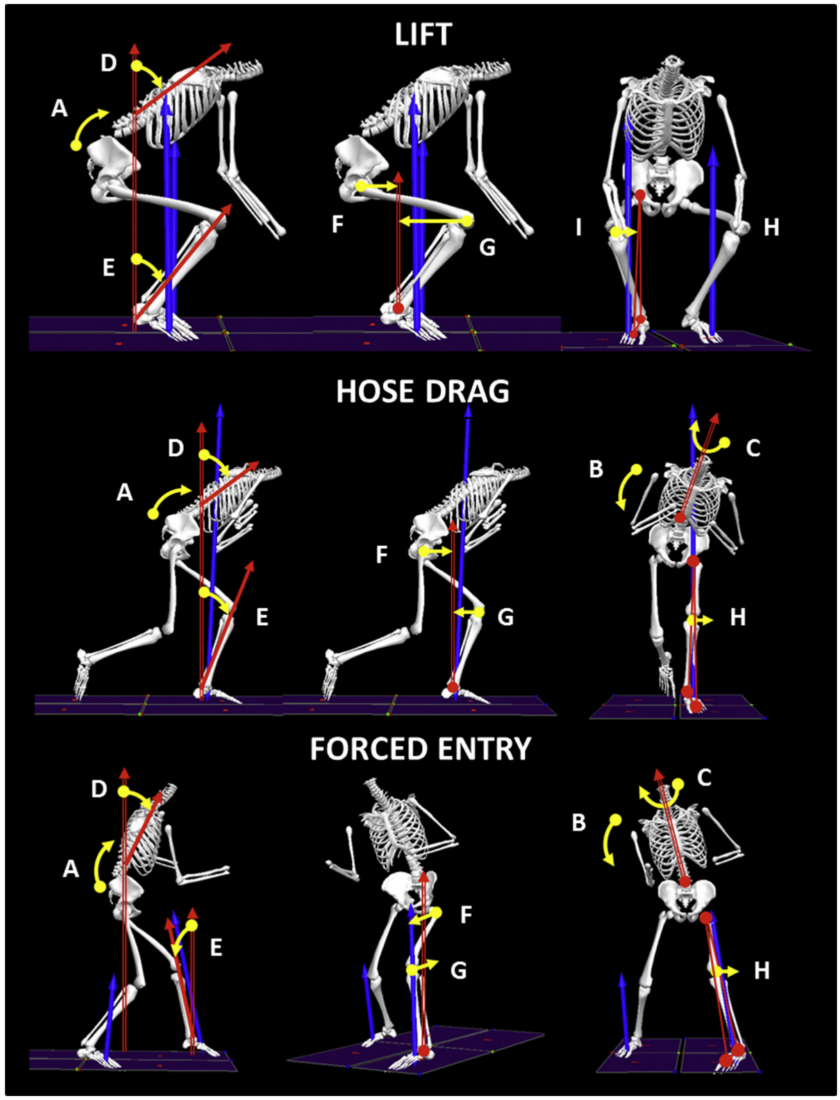


Fig. 1. Participants' movement patterns were characterized with the following variables: (A) spine flexion/extension; (B) spine lateral bend; (C) spine axial twist; (D) trunk angle; (E) shank angle; (F) hip-ankle distance; (G) knee-ankle distance; (H) left knee position; and (I) right knee position.

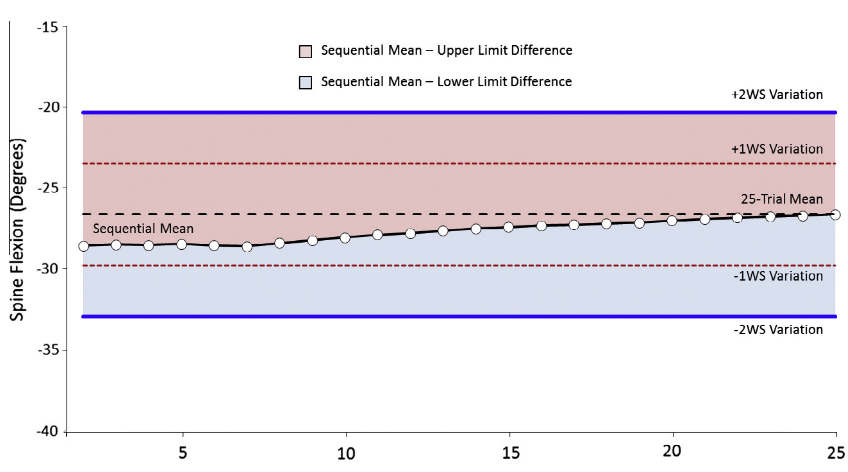


Fig. 2. The sequential mean, 25-trial mean, and 25-trial variation (±1 SD and 2 SD) for one sample variable. The shaded area above the sequential mean reflects the difference between the upper boundary of the group's 25-trial mean (±2 SD). The shaded area below the sequential mean reflects the lower boundary difference. Regression equations were computed to establish a relationship between the magnitudes of these shaded areas, the number of trials included in the sequential mean, and the group's within-subject (WS) variation.

same. The utility of each equation was then evaluated by computing the number of instances across all variables, tasks and testing sessions whereby the computed upper and lower limits extended beyond the measured 25-trial mean + 1 SD and – 1 SD. The number of successful instances was expressed as percentage of the total number of possible variable, task and session computations. For comparative purposes, the utility of a fourth method that relied solely on the group's within-subject variation was also examined. Limits were defined by the following equation: $y = 2z + b$; whereby y and b were the mean and SD of the group's sequential within-subject variation, respectively. For this method, boundary criteria were created using 1, 1.5 and 2 SD.

3. Results

3.1. General tasks (lifting)

For both the light and heavy lifting tasks, *Session* was a significant factor for TRK and HIP ($p < 0.042$). But with the exception of peak HIP during the heavy condition, differences were limited to the mean and the descent phase. *Subject* was a significant factor for every variable investigated ($p < 0.001$), indicating heterogeneity amongst the group.

With regards to the within-subject variation, significant *Session* effects were found for mean and peak SHK and KNE ($p < 0.029$) during the descent phase of the heavy lift and mean and peak TRK ($p < 0.031$) for the light lifting task (Fig. 3). *Subject* was a significant factor for each variable investigated ($p < 0.020$).

Increasing the external load of the lifting task had a significant impact on the magnitude of the within-subject variation for 3 of 21 variables investigated (peak FLX and KNE, and mean TRK during the descent; Fig. 3).

3.2. Specific tasks (firefighting)

No significant between-*Session* differences were found for any of the variables used to characterize the forced entry task (Fig. 4). Only 7 of the 24 hose drag variables were significantly different between sessions. Once again, *Subject* was a significant factor for each variable investigated ($p < 0.001$).

Session had a significant impact on the within-subject variation of six hose drag (max TRK, KNE, LFT and mean FLX, TST, TRK; $p < 0.049$) and two forced entry (min BND and mean HIP) variables (Fig. 3). *Subject* was a significant factor across all variables ($p < 0.001$).

The magnitude of the within-subject variation was significantly different between the hose drag and forced entry tasks for 13 of the 24 variables examined (max FLX, TRK, HIP and KNE, all minimums with the exception of HIP and KNE, and mean HIP, KNE and LFT; Fig. 4).

3.3. Within-subject differences

Averaged across all variables, tasks and testing sessions, the mean and standard deviation of the group's within-subject variation decreased and increased, respectively, as fewer trials were used to compute the sequential mean (Fig. 5). A similar finding was also noted when contrasting the sequential means with the upper and lower limits defined by the 25-trial variation (Fig. 6). A non-linear relationship was noted between the number of trials included in the sequential mean and the magnitude of the difference defined herein as a *bona fide* difference. The following three regression equations were generated based on the abovementioned findings:

(1) $y = -0.0005x^3 + 0.0253x^2 - 0.4491x + 5.0763$ ($r^2 = 0.971$, $p < 0.001$).
(2) $y = 0.0052x^2 - 0.2207x + 4.4325$ ($r^2 = 0.928$, $p < 0.001$).
(3) $y = -0.0796x + 3.7304$ ($r^2 = 0.797$, $p < 0.001$).

In each instance, y would be the magnitude of a *bona fide* difference (i.e. boundary criteria outside of which would be described as an actual change), and x is the number of trials collected. The number of instances across all variables, tasks and testing sessions whereby the 25-trial variation was contained within the boundaries established by the cubic, quadratic and linear regression equations increased as the aggregate scores comprised more trials (Fig. 7). However, each approach achieved a success rate of at least 88% with only three trials. The limits defined by each of the regression equations were able to capture the 25-trial variation in 95% of all instances when five trials were used. In comparison, the 4th method examined that relied exclusively on the within-subject variation had success rates ranging from 74–84% to 85–92% for three and five trials, respectively (Fig. 6). Upper and lower limits were defined successfully in 96% of all instances with 10 trials.

4. Discussion

In general, there were few between-session differences found in the variables chosen to describe each task; however, in every instance, *Subject* was found to be a significant factor. This implies that participants exhibited unique movement behaviors that did not necessarily reflect that of the group or each other. Although this variation has often been perceived as noise that affects the power to detect differences between multiple conditions (van Dieen et al., 2002), it could also be functional (Davids et al., 2003) and reveal important information regarding the task, environment or movement strategies employed by the individuals being studied (Mathiassen et al., 2003). For example, knowing how much performers' spine flexion varies during a heavy lifting task could help to monitor the effects of a short- or long-term intervention on an individual level (Scholes et al., 2012).

Because the sources and roles of movement variability are theoretically and experimentally challenging to interpret and accommodate, respectively, both the between- and within-subject variation are frequently reported by authors exploring the effect of a particular condition or intervention (e.g. Granata et al., 1999; Grills et al., 1994; Kjellberg et al., 1998; Mathiassen et al., 2003; Mirka and Baker, 1996; Scholes et al., 2012; van Dieen et al., 2002). Several metrics have been used to describe this dispersion, although the most widely adopted may be the coefficient of variation (CV) given that it provides a normalized estimate that can be contrasted against other variables and used to make comparisons with earlier work. However, a CV may have little meaning if it is not computed on ratio scale data and thus its utility in helping to define actual changes is limited. For instance, the CVs of the maximum and minimum knee to ankle distance during session one of the forced entry task in this study were 31% and 232%, respectively. This finding suggests that the maximum distance was far more repeatable, but because the mean of the minimum distance was near zero, the CV was not an appropriate descriptor of the variation. As an alternative, the variation of each dependent measure in this study was described by dividing the within-subject variation by that seen between participants. This approach ensured that comparisons could still be made across variables, between tasks, and with future research. Not surprisingly, there was more variation amongst each group than was observed within the individual participants, which lends further support to the notion that in many cases, the interpretation of any findings could be constrained by the heterogeneity of the population studied (Caster and Bates,

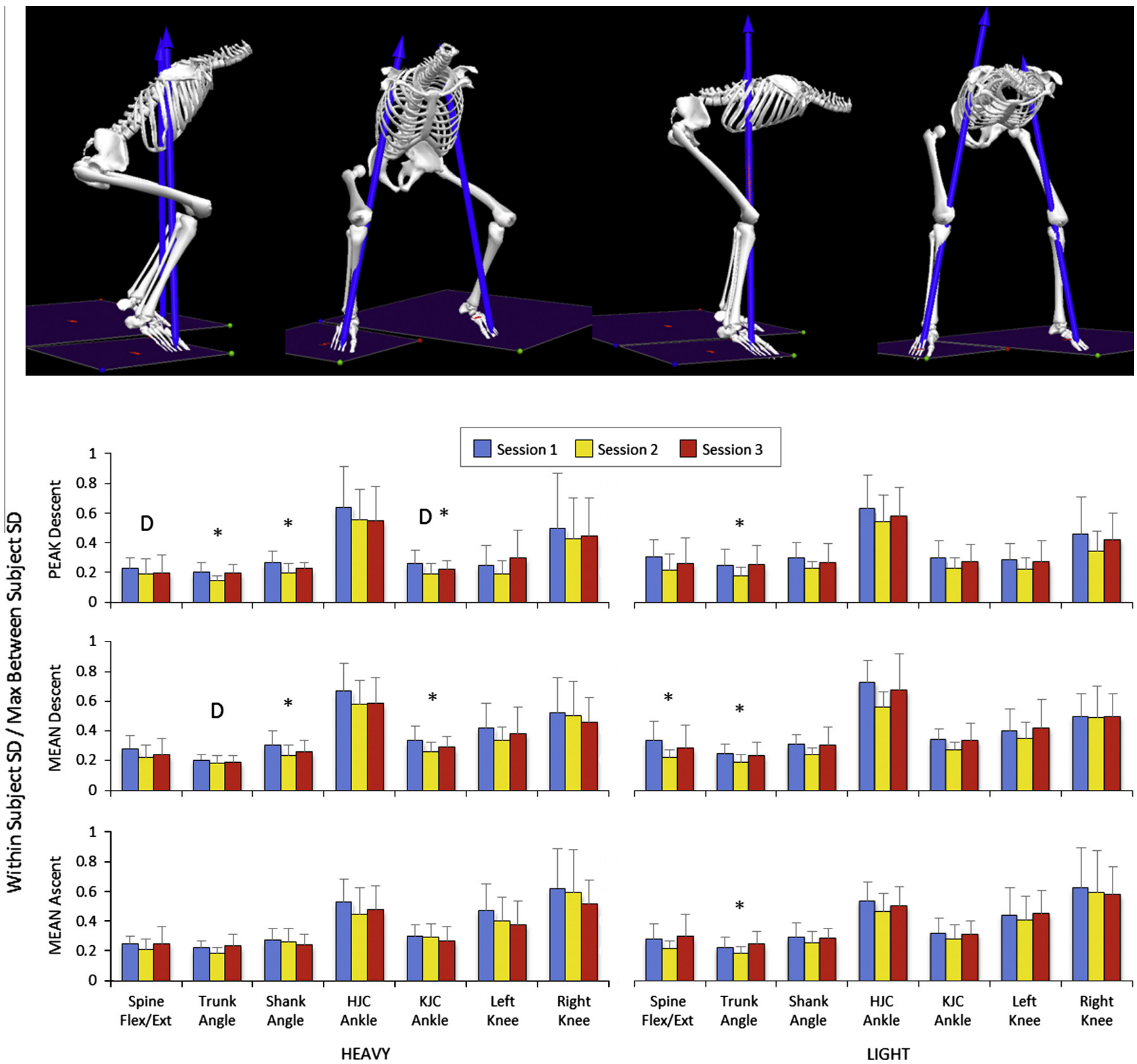


Fig. 3. The 25-trial mean (SD) within-subject variation exhibited during sessions 1, 2 and 3 for the HEAVY and LIGHT lifting tasks. Variation in the peak and mean of the descent phase (unloaded) and the mean of the ascent phase (load in hands) are presented as a function of the maximum between-subject SD observed for a given variable (i.e. that of the peak, mean descent or mean ascent). Significant session effects ($p < 0.05$) are described with an $*$. Instances marked with a *D* denote a significant difference ($p < 0.05$) in the *within-subject variation* observed between the two conditions. Although not shown, *subject* was a significant factor across all variables for both conditions. The model animations depict two unique movement strategies that were used to perform the lifting tasks.

1995; Dufek and Bates, 1990; Dufek et al., 1995; Scholes et al., 2012).

Various statistical analyses have been used to determine the minimum number of trials necessary to achieve a stable estimate of the mean for a range of variables and activities (Bates et al., 1983; DeVita and Bates, 1988; Dunk et al., 2005; Hamill and McNiven, 1990; James et al., 2007; Rodano and Squadrone, 2002; Stuelcken and Sinclair, 2009). As was shown in this study, increasing the number of trials collected for a particular condition can also reduce the standard deviation of the group (van Dieen et al., 2002). But while these approaches have helped to highlight the potential limitations in collecting too few trials and brought attention to the impact of movement variability, they have not necessarily offered a viable solution to deal with the variation that is commonly seen within and between participants across a range of methodological designs, nor have they provided a means to describe participant-specific or actual changes. Slight modifications to a task, condition or the population being tested will likely alter the variation associated with a particular metric, and thus require a different number of trials to achieve a stable estimate of the dispersion. Further, and perhaps more importantly, collecting a large number of trials is often not feasible or appropriate to test hypotheses.

The method proposed in this paper to examine the relevance of within-subject differences is comparable to previous work that has sought to describe clinical differences (e.g. Knutson, 2005) or make meaningful inferences using confidence limits (e.g. Batterham and Hopkins, 2006), but instead used participants' variation to define boundary criteria. As expected, its utility did improve as more trials were used to compute the upper and lower limits; however, using an average of only three repetitions was still able to capture the measured 25-trial variation in 88% of all instances, averaged across all variables and tasks (using five trials increased the success rate to 95%). Although these findings are encouraging, further work is needed to examine the merit of the equations proposed using a different set of variables, tasks and participants, before it can be stated that the boundary criteria do in fact reflect actual differences.

In summary, the findings of this study show that while movement is variable between individuals who perform the same tasks, it may be possible to establish boundary criteria for each performer

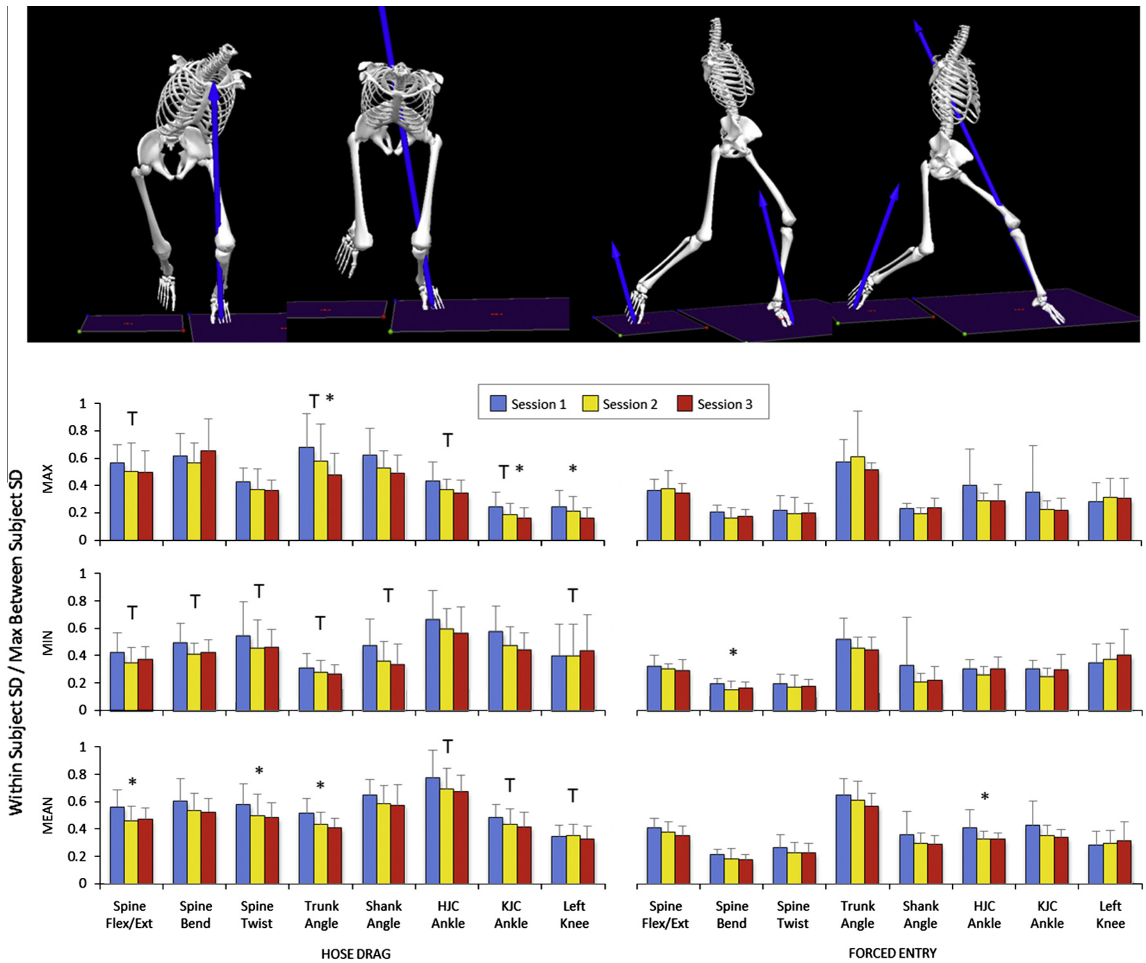


Fig. 4. The 25-trial mean (SD) within-subject variation exhibited during sessions 1, 2 and 3 for the simulated HOSE DRAG and FORCED ENTRY tasks. Variation in the max, min and mean of each task is presented as a function of the maximum between-subject SD observed for a given variable (i.e. that of the max, min or mean). Significant session effects ($p < 0.05$) are described with an $*$. Instances marked with a *T* denote a significant difference ($p < 0.05$) in the *within-subject variation* observed between the two tasks. Although not shown, *subject* was a significant factor across all variables for both tasks. The model animations depict two unique movement strategies that were used to perform the hose drag and forced entry tasks.

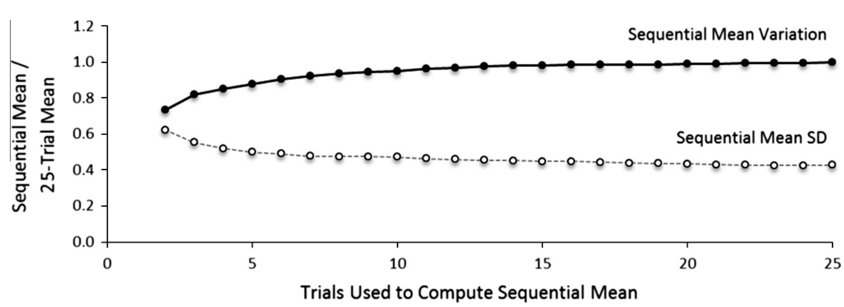


Fig. 5. The mean and standard deviation of participants' within-subject variation across all variables, tasks and testing sessions. The sequential variation was normalized by that measured over 25 trials so that a general relationship could be established for all variables.

outside of which would reflect a *bona fide* change in motion. Several factors including perceived risk, previous experience, physical capacity (structural and functional), and environmental conditions can interact to constrain task performance (Davids et al., 2003), which is why efforts must be made to ensure that the experimental protocols, instrumentation and data analyses are appropriate to test the stated hypotheses. If these factors are not adequately accounted for (theoretically and experimentally), the understanding of how and why individuals' adapt their movement behavior in response to various interventions will be compromised. Performers may adapt their movement strategies if asked to complete multiple trials of the same task (e.g. due to motor learning- and/or fatigue-related processes), but based on the findings of this investigation, the magnitude of this dispersion appears to be much smaller than what would be observed amongst a group – at least in the tasks studied here. When an individual exhibits a change in their movement behavior beyond what would be considered "typical", it should arguably be described as an "actual" difference. Although much more evidence is needed to substantiate the method proposed in this study, it may offer an

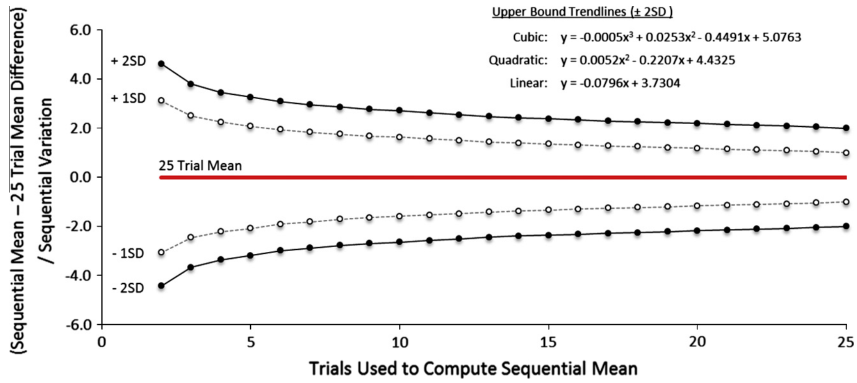


Fig. 6. The differences between each individual's sequential means and the upper and lower limits of their biological variation, as defined by the 25-trial mean ± 1 SD and 2 SD. In each instance, the difference scores were normalized by the group's sequential within-subject variation such that boundary criteria could be generalized to any number of trials. The results across all participants, variables, tasks and testing sessions were collapsed and used to create three generalizable regression equations (i.e. cubic, quadratic, linear) that could help to define limits for within-subject differences.

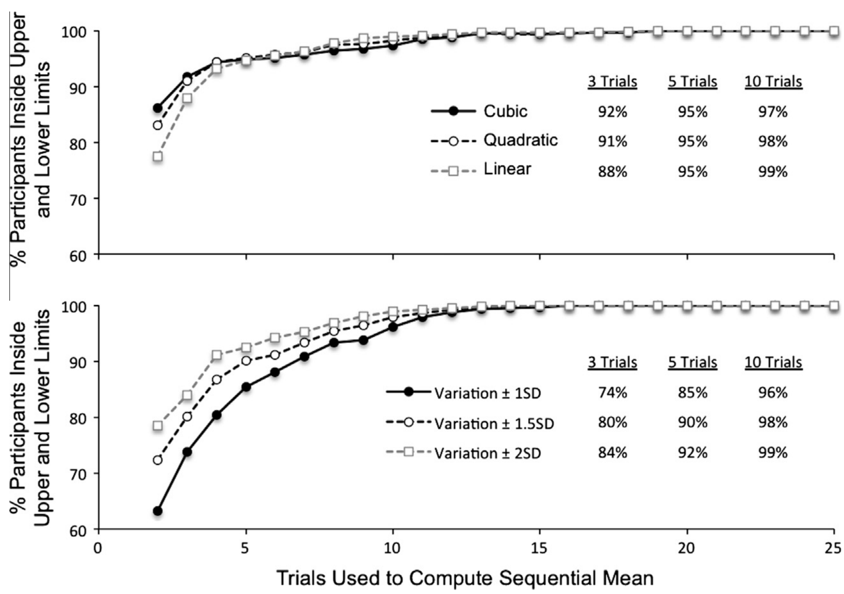


Fig. 7. The instances across all variables, tasks and testing session (expressed as a percentage of the total number possible) whereby the computed upper and lower limits, as defined by the regression equations and/or within-subject variation, extended beyond the measured 25-trial mean +1 SD and −1 SD. The success rate for 3, 5 and 10 trials is also listed.

effective means to explore changes in an individual's behavior by exploiting their own between-trial variation.

5. Conflict of interest

The authors wish to declare that they have no conflicts of interest.

Acknowledgements

We would like to extend our gratitude to each member of Waterloo and Kitchener Fire Departments for their commitment to this work. Funding for this project was provided by the Centre of Research Expertise for the Prevention of Musculoskeletal Disorders (CRE-MSD) and the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr. Jack Callaghan is also supported by the Canada Research Chair in Spine Biomechanics and Injury Prevention.

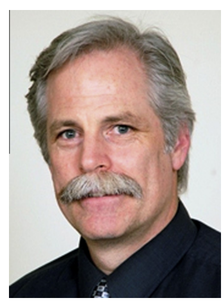
References

Bates BT, Osternig LR, Sawhill JA, James SL. An assessment of subject variability, subject-shoe interaction, and the evaluation of running shoes using ground reaction force data. J Biomech 1983;16(3):181–91.
Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. Int J Sports Physiol Perform 2006;1(1):50–7.
Begon M, Monnet T, Lacouture P. Effects of movement for estimating the hip joint centre. Gait Posture 2007;25(3):353–9.
Callaghan JP, McGill SM. Intervertebral disc herniation: studies on a porcine model exposed to highly repetitive flexion/extension motion with compressive force. Clin Biomech 2001;16(1):28–37.
Caster BL, Bates BT. The assessment of mechanical and neuromuscular response strategies during landing. Med Sci Sports Exerc 1995;27(5):736–44.
Cole GK, Nigg BM, Ronsky JL, Yeardon MR. Application of the joint coordinate system to three-dimensional joint attitude and movement representation: a standardization proposal. J Biomech Eng 1993;115(4A):344–9.
Davids K, Glazier P, Araujo D, Bartlett R. Movement systems as dynamical systems: the functional role of variability and its implications for sports medicine. Sports Med 2003;33(4):245–60.
DeVita P, Bates BT. Intraday reliability of ground reaction force data. Hum Mov Sci 1988;7(1):73–85.
Dufek JS, Bates BT. The evaluation and prediction of impact forces during landings. Med Sci Sports Exerc 1990;22(3):370–7.
Dufek JS, Bates BT, Stergiou N, James CR. Interactive effects between group and single-subject repones patterns. Hum Mov Sci 1995;14(3):301–23.

Dunk NM, Keown KJ, Andrews DM, Callaghan JP. Task variability and extrapolated cumulative low back loads. Occup Ergon 2005;5(3):149–59.
Frost DM, Gooyers CE, Campbell TL, McGill SM, Callaghan JP. Might the interpretation of between-day changes in joint angles, forces and moments be influenced by variation in the link segment model? In: Proceeding of the: 17th biennial conference of the Canadian society for biomechanics. Vancouver, British Columbia; 2012.
Granata KP, Marras WS, Davis KG. Variation in spinal load and trunk dynamics during repeated lifting exertions. Clin Biomech 1999;14:367–75.
Grills PM, Mientjes M, Wells R, Norman R, Neumann P, Frank J, et al. The reliability and variability of low back physical exposure measures. In: Proceeding of the: 12th triennial congress of the international ergonomics association. Toronto, Ontario; 1994.
Hamill J, McNiven SL. Reliability of selected ground reaction force parameters during walking. Hum Mov Sci 1990;9(2):117–31.
Harbili E, Alptekin A. Comparative kinematic analysis of the snatch lifts in elite male adolescent weightlifters. J Sports Sci Med 2014;13(2):417–22.
Hewett TE, Myer GD, Ford KR, Heidt RS, Colosimo AJ, McLean SG, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. Am J Sports Med 2005;33(4):492–501.
Hopkins W. Measures and reliability in sports medicine and science. Sports Med 2000;30(1):1–15.
James CR, Bates BT. Experimental and statistical design issues in human movement research. Meas Phys Educ Exerc Sci 1997;1(1):55–69.
James CR, Herman JA, Dufek JS, Bates BT. Number of trials necessary to achieve performance stability of selected ground reaction force variables during landing. J Sports Sci Med 2007;6(1):126–34.
Kjellberg K, Lindbeck L, Hagberg M. Method and performance: two elements of work technique. Ergonomics 1998;41(6):798–816.
Knutson GA. Anatomic and functional leg-length inequality: A review and recommendation for clinical decision-making. Part I, anatomic leg-length inequality: prevalence, magnitude, effects and clinical significance. Chiropractic Osteopathy 2005;13(11).
Marras WS, Lavender SA, Leurgans SE, Rajulu SL, Allread WG, Fathallah FA, et al. The role of dynamic three-dimensional trunk motion in occupationally-related low back disorders. The effects of workplace factors, trunk position, and trunk motion characteristics on risk of injury. Spine 1993;18(5):617–28.
Marshall LW, McGill SM. The role of axial torque in disc herniation. Clin Biomech 2010;25(1):6–9.
Mathiassen SE, Moller T, Forsman M. Variability in mechanical exposure within and between individuals performing a highly constrained industrial work task. Ergonomics 2003;46(8):800–24.
Mirka GA, Baker A. An investigation of the variability in human performance during sagittally symmetric lifting tasks. IIE Trans 1996;28(9):745–52.
Morriss C, Bartlett RM, Fowler N. Biomechanical analysis of the men's javelin throw at the 1995 world championships in athletics. New Stud Athl 1997;12(2):31–41.
Punnett L, Fine LJ, Keyserling WM, Herrin GD, Chaffin DB. Back disorders and nonneutral trunk postures of automobile assembly workers. Scand J Work Environ Health 1991;17(5):337–46.
Rodano R, Squadrone R. Stability of selected lower limb joint kinetic parameters during vertical jump. J Appl Biomech 2002;18(1):83–9.
Scholes CJ, McDonald MD, Parker AW. Single-subject analysis reveals variation in knee mechanics during step landing. J Biomech 2012;45(12):2074–8.
Schwartz MH, Rozumalski A. A new method for estimating joint parameters from motion data. J Biomech 2005;38(1):107–16.
Stuelcken MC, Sinclair PJ. A pilot study of the front foot ground reaction forces in elite female fast bowlers. J Sci Med Sport 2009;12(2):258–61.
van Dieen JH, Hoozemans MJM, van der Beek AJ, Mullender M. Precision of estimates of mean and peak spinal loads in lifting. J Biomech 2002;35(7):979–82.

David Frost is an Assistant Professor in the Faculty of Kinesiology and Physical Education at the University of Toronto. His research is focused on the development of education, coaching, and exercise strategies to improve the physical capacity of (occupational) athletes.

Tyson Beach is an Assistant Professor in the Faculty of Kinesiology and Physical Education at the University of Toronto where he also holds an appointment in the Graduate Department of Exercise Sciences. In addition to his teaching and mentoring roles, he directs the Musculoskeletal Biomechanics and Injury Prevention Laboratory (http://physical.utoronto.ca/biomechanics.aspx). An overarching aim of his research program is to devise sustainable education- and exercise-based strategies to enhance and maintain the capacity of the musculoskeletal system to withstand physical demands of work and sport.



Stuart McGill is a Professor of Spine Biomechanics at the University of Waterloo. He works to find solutions for challenging back patients and athletes.

Jack Callaghan is a Professor in the Department of Kinesiology at the University of Waterloo, the Canada Research Chair in Spine Biomechanics and Injury and Prevention, and the Associate Director, Research for the Centre of Research Expertise for the Prevention of Musculoskeletal Disorders (CRE-MSD). He uses a combination of in-vivo and in-vitro approaches to examine the biological responses to cumulative loading exposure from both pain generating and tissue altering/injuring perspectives.